

A Machine Learning Strategy for Locomotion Classification and Parameter Estimation using Fusion of Wearable Sensors

Jonathan Camargo^{1*}, Will Flanagan^{1*}, Noel Csomay-Shanklin², Bharat Kanwar¹, Aaron Young¹

Abstract—The accurate classification of ambulation modes and estimation of walking parameters is a challenging problem that is key to many applications. Knowledge of the user’s state can enable rehabilitative devices to adapt to changing conditions, while in a clinical setting it can provide physicians with more detailed patient activity information. This study describes the development and optimization process of a combined locomotion mode classifier and environmental parameter estimator using machine learning and wearable sensors. A detailed analysis of the best sensor types and placements for each problem is also presented to provide device designers with information on which sensors to prioritize for their application. For this study, 15 able-bodied subjects were unilaterally instrumented with inertial measurement unit, goniometer, and electromyography sensors and data were collected for extensive ranges of level-ground, ramp, and stair walking conditions. The proposed system classifies steady state ambulation modes with 99% accuracy and ambulation mode transitions with 96% accuracy, along with estimating ramp incline within 1.25 degrees, stair height within 1.29 centimeters, and walking speed within 0.04 meters per second. Mechanical sensors (inertial measurement units, goniometers) are found to be most important for classification, while goniometers dominate ramp incline and stair height estimation, and speed estimation is performed largely with a single inertial measurement unit. The feature tables and Matlab code to replicate the study are published as supplemental materials.

Index Terms—intent recognition, activity monitoring, wearable sensors, sensor fusion.

I. INTRODUCTION

As of 2017, over 20 million people reported living with an ambulatory disability in the US alone [1], a number that will only increase with the growing elderly population. Aside from aging, these disabilities can be the result of spinal cord injuries, strokes, or disease. In each case, proper rehabilitation therapy is a very important process for regaining mobility, independence, and reducing the risk of further health complications [2]. As technology becomes more sophisticated and portable, the application of *wearable sensors* – inertial measurement unit (IMU), goniometer (GON), and electromyography (EMG) – to rehabilitation has become a prevalent field of research.

These sensors have been used to provide additional information to clinicians [3] as well as improve the control of rehabilitative devices such as prostheses [4], orthoses, and exoskeletons [5]. The machine learning strategy presented in this study focuses on the latter application, although it could be adapted for patient activity monitoring.

Powered rehabilitative devices are designed to replicate or augment biological torques at the ankle, knee and hip joints using advanced sensing and control modes. Currently, many research devices operate using variable power and damping parameters to assist with walking in various community *ambulation modes* – level ground, ramp, and stair walking [6][7][8]. These control parameters are manually tuned for each ambulation mode based on patient comfort and performance feedback [9], a time-consuming process that is conducted during initial testing and normally under a single set of *environmental parameters* – walking speed, ramp incline, or stair height. More recent studies show promise of using EMG to predict human intention, online mode classification algorithms, and model-based torque control leveraging kinematic and force data [10], [11], [12], although these systems largely assume a static environment. Because gait kinematics vary with each environmental parameter [13][14][15], using static control parameters results in a device unable to adapt to changing environments. This adaptability is also important for monitoring rehabilitation, as walking speed is a common outcome measure [16], and the use of stair [17] and ramp [18] exercises improves muscle strength, balance, and walking speed. For adaptation to occur within a device’s control architecture without user intervention, however, both an accurate estimate of the relevant environmental parameter and classification of the user’s ambulation mode is required.

Locomotion mode classifiers have been widely analyzed within the field of rehabilitative robotics, with the common models being linear discriminate analysis (LDA) and support vector machines (SVM) [19], [20], [4]. Other models like kNN, k-means and Gaussian Mixture Models have also been studied, but without significant improvement in comparison to LDA [21]. A common technique to improve classification accuracy beyond the base model is the addition of time-history information. This is typically implemented as a majority vote filter [4], although studies have shown that a dynamic Bayesian network (DBN) model – a type of hidden Markov model (HMM) which uses a probabilistic approach to incorporate information from previous and current decisions based on the model’s confidence in each decision – performs better than an

¹George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA. {jon-cama, wflan, bkanwar3}@gatech.edu, aaron.young@me.gatech.edu

²Department of Control and Dynamical Systems, California Institute of Technology, Pasadena, CA 91125 USA. noelcs@caltech.edu

*These authors contributed equally to this work

Copyright (c) 2021 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieee.org.

LDA with majority vote [22]. Other advanced models could be applied to this problem as well, such as Long Short Term Memory (LSTM) neural networks, however these were not investigated due to their increased complexity. In this study, the performance between LDA and SVM models was compared along with the effect of time-history information implemented as a DBN.

While locomotion classification has been heavily analyzed, research into the estimation of continuous environmental parameters using wearable sensors is comparatively sparse. The most common methods for estimation are the integration of IMU data [23], kinematic models [24], and machine learning (ML) approaches [25][26]. Although integration techniques are easily applied to speed and slope estimation, accurate height estimation is more prone to error due to the required double integration. Of the three estimation methods, kinematic model-based methods are the most accurate for speed estimation [24]; however, the creation of kinematic models requires detailed analysis of each ambulation mode. ML methods for walking parameter estimation have recently received attention because of the ability to frame the input-output nature of wearable sensors and walking parameters as a regression problem. This relationship lends itself well to a generalized approach for estimation over various ambulation modes and is easily expanded to incorporate information from many sensors. The application of multi-sensor fusion to this problem is important as an early evaluation of neural networks for regression from our research group showed that the addition of EMG information along with multiple IMU and GON sensors improved estimation in a robotic hip exoskeleton [27]. Along this thread, we aim to further extend the ML approach of sensor fusion by expanding the number of sensors and providing the first instance of stair height estimation. Two common regression models were compared, a feed-forward neural network (NN) and a regression SVM. The application of a Kalman filter to each model output was also studied, as this filtering method is an efficient way of incorporating time-history into discrete estimations [28], and was previously implemented in integration-based speed estimation to reduce noise [29].

Although the use of many sensors and features for ML models tends to increase performance, the addition of inputs greatly increases the complexity of the model. In order to reduce the input space of the model, feature selection algorithms are normally applied. While these techniques are an active field of research and many methods could have been explored, sequential forward feature selection was chosen due to its simplicity and low error to time-cost ratio compared to exhaustive search methods [30]. To further refine the model, additional optimization steps were performed for both classification and regression to determine the best combination of model parameters for these specific applications. Finally, as previous studies have shown that the fusion of mechanical and EMG sensors improve classification [4], [31] and regression [27], a more detailed analysis of the relative importance of each sensor type along with sensor location was conducted. This should inform designers on which sensors are most needed when implementing this system, and which additional

sensors would provide the greatest benefit to existing systems.

In this paper, we present a machine learning strategy for determining a user's ambulation state through a locomotion mode classifier feeding into a continuous parameter estimator for level ground, ramp, and stair modes. The system was applied over a very extensive range of environmental parameters (6 ramp inclines, 4 stair heights, and 28 walking speeds) to increase robustness and better represent the continuous nature of real-world environments. The process of constructing this system is detailed through the feature selection, parameter sweeping, model selection, and model optimization steps. We also aim to provide designers of new and existing systems with insight on which sensors are most important for each model so that the effort of incorporating wearable sensors is spent most effectively. Previous studies have shown that EMG information aids in classification, since these signals relate to the user's intent. In contrast, we hypothesize that parameter estimation can be performed largely through mechanical sensors, as these sensors indicate the geometry and kinematics of each mode. The primary contributions of this work include: 1) The development and rigorous optimization of an ML approach for locomotion classification and environmental parameter regression, 2) A detailed analysis of best sensor types and placements for four ML problems: ambulation mode classification, walking speed estimation, ramp incline estimation, and stair height estimation, and 3) An extensive open-source, multi-terrain feature dataset.

II. METHODS

A. Experimental Protocol

15 healthy, able-bodied adults (age 21 ± 3.4 yr, height 1.70 ± 0.07 m, weight 68.3 ± 10.83 kg) participated in this study after providing informed consent. This experiment was approved by Georgia Institute of Technology IRB under protocol No. H17240.

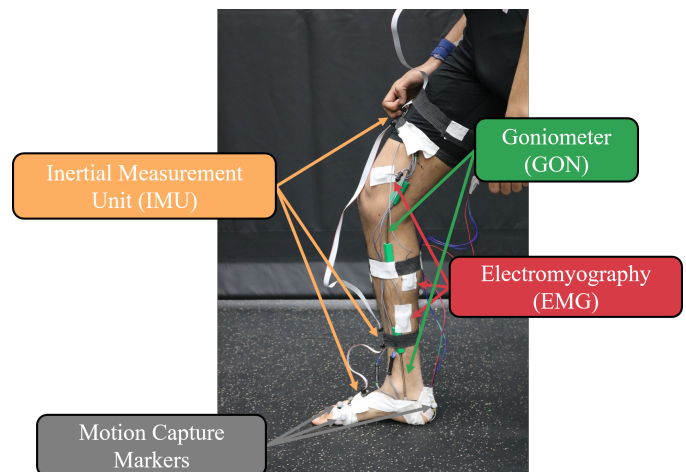


Fig. 1: A subset of the full sensor set used for data collection. A combination of inertial, joint angle, muscle activity, and global position sensors were used to characterize the subjects' walking.

1) *Instrumentation*: Subjects were unilaterally instrumented with 11 EMG sensors, 3 GON, 4 IMU, and 32 motion capture markers following the Plugin-gait/Cleveland Clinic marker set [32] (Figure 1). For EMG, sensors were placed to measure the gastrocnemius medialis, tibialis anterior, soleus, vastus medialis, vastus lateralis, rectus femoris, biceps femoris, semitendinosus, gluteus medius, and right external oblique using SENIAM recommendations and palpation. All EMG locations were shaved and cleaned with alcohol prior to placement. For GON, angles for both frontal and sagittal planes were measured for the hip and ankle joints, while only the sagittal plane was measured for the knee. The IMUs were placed on each of the four segments: trunk, thigh, shank, and foot. The EMG, GON, and IMU sensors were sampled at 1000, 50, and 250 Hz, respectively, with the IMU and GON sensors being upsampled to 1000 Hz after data collection was complete.

2) *Ambulation Modes*: Every subject completed trials on four different modes: treadmill, level ground, ramps, and stairs (Figure 2). A video of trials for each ambulation mode is provided in the supplemental materials.

For level ground walking, subjects performed counterclockwise and clockwise circuits at 3 self-selected speeds, where subjects were instructed to walk at slower than their preferred pace, preferred, and faster than their preferred pace. Based on motion capture data, walking velocities were found to be 0.88 ± 0.19 m/s for slow self-selected speed, 1.17 ± 0.21 m/s for preferred, and 1.45 ± 0.27 m/s for fast. The U-shaped circuit started with a stand-to-walk transition, then a straight walking segment followed by two consecutive 90 degree turns, a second straight walking segment, and ended with a walk-to-stand transition. 5 trials were recorded for each speed and direction for a total of 30 walking trials.

Data were also collected for 6 different ramp inclines (5.2, 7.8, 9.2, 11, 12.4, 18 deg) and 4 different stair heights (10.1, 12.7, 15.2, 17.8 cm). These activities were performed using a 5 meter long ramp and 6 step staircase connected to an adjustable-height platform. For each ramp/stair trial, subjects started from standing, approached the ramp/stair, ascended, and came to a full stop on top of the platform. Subjects then turned in place and repeated the same process for the descent. Additionally, subjects were instructed to complete a set of trials starting with their instrumented leg and another set starting with their non-instrumented leg at each level. 5 trials were recorded for each starting leg for a total of 60 ramp and 40 stair trials.

Treadmill walking data were collected at 28 different speeds over 7 trials (4 speeds per trial), ranging from 0.5 m/s to 1.85 m/s in 0.05 m/s increments. Each trial started from standing, with steady state walking data for 30 seconds at each speed and 5 seconds to accelerate between speeds. These trials produced additional walking data at a high resolution of speeds.

Ambulation mode labels were created based on motion capture data, with transition steps identified as the last toe-off on the previous mode to the first toe-off on the next mode for the instrumented leg, similar to [33], with no differentiation between the instrumented leg being leading or trailing. To

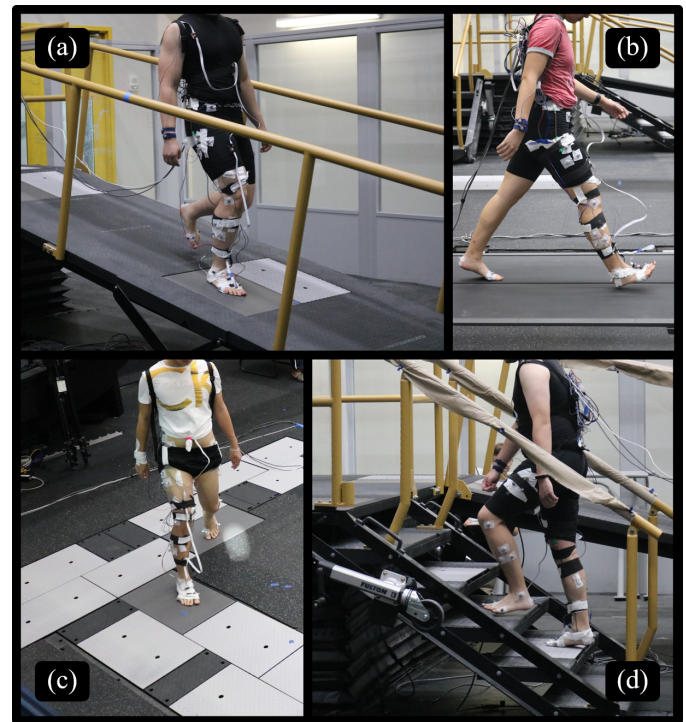


Fig. 2: Subjects performing the tested ambulation modes: (a) ramps, (b) treadmill, (c) level ground, (d) stairs.

simplify generalizations between steady state and transitional steps, the gait cycle was defined as toe-off to toe-off and gait phase was linearly interpolated between these two points. Transitions were not labeled as separate classes, instead the mode labels changed at 50% gait phase during the transition step, which roughly corresponds to heel strike. Detecting gait events with motion-capture is not the only approach however, as other studies have used inertial sensors, ground reaction forces and kinematic models [34], [35], [36], [37]. As discussed in the next section, the proposed strategy is modular such that any gait phase estimator can be used in place of our motion-capture-based method.

B. Machine Learning

The proposed approach uses a cascaded style of machine learning with a classifier layer feeding into a regressor layer. The system operates as described in Figure 3. Wearable sensor data are fed into a feature extractor, where different features are computed based on the sensor type and the user's gait phase. The features are first used by the intent classifier to determine the ambulation mode. The chosen mode then determines the environmental parameter to estimate: if the user's mode was classified as level ground walking, walking speed was estimated, while a classification of ramp ascent or descent triggered incline estimation, and classification of stair ascent or descent triggered stair height estimation. The estimated values are subsequently smoothed by a Kalman Filter. To optimize this strategy, multiple steps (described in the following sections) were performed to select the best features, choose the best model for classification and regression, and determine if the addition of time-history would significantly

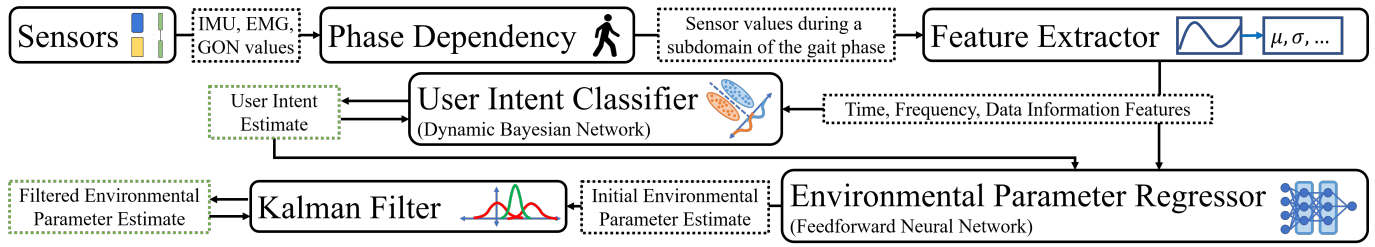


Fig. 3: The general strategy used for learning walking parameters. Wearable sensor data are fed into a feature extractor, where different features are extracted depending on the sensor type and the user’s gait phase (phase dependency). The features are first used by the intent classifier, to determine the user’s intent and subsequently which environmental parameter to estimate. The estimated values are then filtered by a Kalman Filter. The final outputs of both the User Intent Classifier and the Kalman Filter – highlighted in green – are then fed back into the respective blocks as a prior for the next step.

decrease the error. Each model was subject dependent, i.e., the training and testing data for a model was from the same subject. Features were selected based on the average errors across all subjects, however, such that all subjects used the same feature set. Models for each subject were evaluated by a k-fold leave-one-out cross-validation strategy based on trials, such that for each fold i , the i^{th} trial was used for testing and all other trials were used for training. This was done to maintain the time-series nature of the data and prevent overfitting. For classification, models were trained on the combined data from 4 of the 5 trials at each condition of ramp, stairs, and level ground walking, and were tested on the remaining trial at each condition of these ambulation modes. Regression models for each mode were trained using all but one trial at each condition of that mode and tested using the remaining trial at each condition. In the case of the ramp incline or stair height regressors, models were trained on 4 of the 5 trials at each incline or height. For walking speed estimation, models were trained on 6 of the 7 trials of the treadmill speeds. Each subject’s reported errors were the averaged results from all the testing sets. These errors are presented as the proportion of incorrect decisions for classification and mean absolute error for parameter estimation. Classification error was further split into transitional and steady state errors based on whether the classification was during a transitional or steady state step. All deviations on errors represent a 99% confidence interval.

1) *Feature Extraction*: A window-based feature extraction method was run on the up-sampled data with different features for each sensor type. The mean absolute value, standard deviation, minimum, maximum, and last values in each window were extracted for mechanical sensors, while the mean absolute value, waveform length, slope sign changes, zero crossings, and 6th order autoregressive coefficients were extracted for EMG sensors [31]. This produced a total of 245 features. In classification, extraction occurred at specific gait phase percentages, termed *classification points*, that were equally spaced over the gait cycle. For regression, feature extraction occurred every 50 ms.

2) *Model Parameters*: Using the approach proposed by Huang et al., separate classifiers and regressors were trained for different points of the gait cycle to create model dependency with respect to the gait phase [20], also referred to as *phase dependency*. Besides input types, several parameters

were used to characterize each model. For classification, these were the number of classification points per gait cycle and the window size for feature extraction at each classification point. A separate model instance was used for each classification point. For regression, the model parameters were window size and number of *phase bins* – an interval of the gait cycle for which a separate model instance was trained – per gait cycle. Due to feature extraction for regression occurring every 50ms, the model instance used was determined by the phase bin in which the feature extraction occurred. The gait cycle was always segmented into phase bins of equal length. Initial models to start the analysis were: an LDA model with 4 classification points using a 400ms window for each, and a NN with 1 hidden layer of 20 nodes estimating at 50ms intervals using 200ms windows and 2 phase bins.

3) *Initial Feature Selection*: To reduce the complexity of each of the models, a subset of features was selected through sequential forward selection methods using the *average error*, i.e. the average of the transitional and steady state errors. At each step of forward feature selection, a new feature is added to the subset of selected features by maximizing the performance of the model with the incremental inclusion of one feature from the set of all possible features that have not been selected (remaining features). For classification, every remaining feature was evaluated at each step. For regression, due to the higher computational cost of the NN compared to the LDA, only a smaller subset of the remaining features was tested at each step. This is termed “restricted forward selection” [30]. In this scheme, the first step is to rank all the features based on the estimation error of every individual feature as the lone input to the NN. The best one is added to the selected features. The next step selects another feature by evaluating 2-feature subsets consisting of the winner of the previous round and any other feature from the top $N/2$ features in the ranking, where N corresponds to the total number of features (245). The process continues and in each i^{th} subsequent step a new feature is selected amongst the top best N/i remaining in the ranking. The value of N/i was limited to a minimum of 5 features. For the initial selection, the first 100 features were chosen for each classification location and the first 20 features were chosen for each environmental parameter estimator. These cutoffs were found to be past the elbow on the feature selection curves during preliminary

analysis, which is the point where the change in error caused by the addition of more features starts to level off. A more extensive selection process is performed later.

4) *Parameter Sweep*: To determine the parameters for each of the models, parameter sweeps were performed using the initially selected features. For classification, different window sizes (50, 100, 200, 300, 400, 500 ms) were tested at each classification point, and the window size that resulted in the least average error was chosen for each location. For regression, the combination of window size (same as classification) and number of phase bins per gait cycle (1, 2, 4, 5, 10) that resulted in the least average error was chosen for each ambulation mode.

5) *Final Feature Selection*: A final selection process was performed using the chosen model parameters. For classification, feature selection was run until the total number of available features was added. For regression, the process stopped when the best 50 features were selected. In both cases, the final feature set was determined as the number of features required to achieve 99% of the *reduction in error*, which was the difference between the initial error and the minimum observed error.

6) *Model Selection and Filtering*: For classification, the initial model was compared to both a linear SVM and a DBN. The SVM was composed of binary learners in a one-vs.-one design with linear kernel and no hyper-parameter optimization. The DBN was constructed as described by Young et al. [22] with the likelihood probability given as the output of an LDA. The transition matrix was initialized using the relative frequencies of each class and transition in the full dataset (combining all subjects) and further optimized using gradient descent. Similarly, for regression, the initial model was compared to a regression SVM. Additionally, an adaptive version of a Kalman filter was applied to both the NN and regression SVM to construct the filtered models. This scheme considers a univariate system with the regression output as the measurement and the measurement noise characterized by the standard deviation of the past regression outputs. The system model was assumed stationary with zero mean and nonzero process noise. The Kalman filter correction was used to estimate the state of the system i.e. walking speed, stair height or ramp inclination. In order to make the filter robust to steady state and transitional behavior, separate process noises for steady state and transitions were chosen along with a noise threshold to switch between them. These values were optimized for each regression model and ambulation mode using `fmincon`, a constrained minimizer in Matlab. In practice, if the standard deviation of the previous 20 model outputs exceeded the noise threshold, then the transitional process noise was used. This scheme allowed the filter to better adapt to changes in output that are due to changes in ambulation condition compared to a standard Kalman filter.

7) *Model Optimization*: After the best models are selected, model-specific parameter sweeps were performed to determine if a different model configuration resulted in significant decreases in error compared to the initial configuration. For classification this parameter was the number of classification points per gait cycle (2, 4, 6, 8), as the change in time between

decisions would affect the DBN output. In the case of the NN model, the number of hidden layers (1, 2) and nodes per hidden layer (10, 20, 30, 40) were swept.

C. Sensor Importance

Using the optimized models and the final selected feature set, both the importance of each sensor type (EMG, IMU, GON) and the sensor locations within a type (e.g. hip, knee, ankle for GON) were investigated. To determine the importance of a sensor type, all models were evaluated after removing the features belonging to that type. After a baseline re-evaluation of the models without the sensors of a specific type, the importance of each sensor location within that type was determined through the relative decrease in error during the forward selection of sensor location. Each step re-introduced the features belonging to a single sensor location until all features of that type were restored.

1) *Statistics*: For every section, a statistical analysis was performed to determine if any given condition of an independent variable resulted in a significant change in error. Within subjects, we computed the average error of all the cross-validation testing sets at each condition. Then we performed a one-way repeated measures ANOVA with Bonferroni PostHoc analysis, testing if there was a difference in error for pairwise comparisons of the conditions with a significance level of $\alpha = 0.01$. For *Model Selection and Filtering*, the independent variable was the type of model. For *Model Optimization*, the independent variable was the number of hidden nodes and layers for NN, and the number of classification points per gait cycle. For *Sensor Importance* the independent variable was the type of sensor excluded.

III. RESULTS

A. Parameter Sweep

The parameter sweep for classification found that the LDA preferred longer window sizes, with 500ms having the lowest error for the two classification points below 50% gait phase (during swing phase) and 400ms for the classification points above 50% (stance phase). Ramp and stair estimation followed a similar trend for window size, with 500ms resulting in the lowest error for all phase bins. Speed estimation differed however, as a minimum occurred at the 300ms window size. The number of phase bins was different for all regression problems, with 5 chosen for ramp estimation, 4 for stairs, and 2 for speed. Full parameter sweep results can be found in the supplemental materials (Figure S1).

B. Final Feature Selection

The feature selection curves for classification and regression are shown in Figure 4. For classification, decisions at 85% gait phase required the fewest number of features (71 features) to obtain 99% total reduction in error (99% of difference between initial and minimum observed error). 10% gait phase was next (88), followed by 60% (126) and 35% (149). For regression, stair height estimation selected the fewest features (10), followed by ramp incline estimation (13), while walking speed required far more (42).

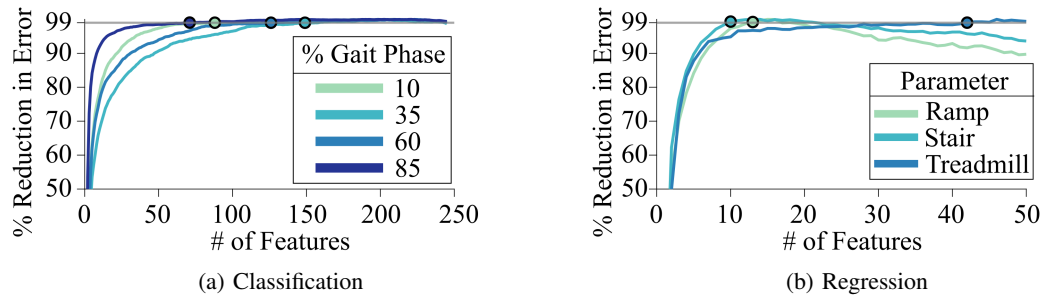


Fig. 4: Feature selection curves for (a) classification and (b) regression. Markers indicate the 99% reduction cutoff. Classification at 35 and 60% gait phase required more features than 10 and 85%. Regression used fewer features than classification.

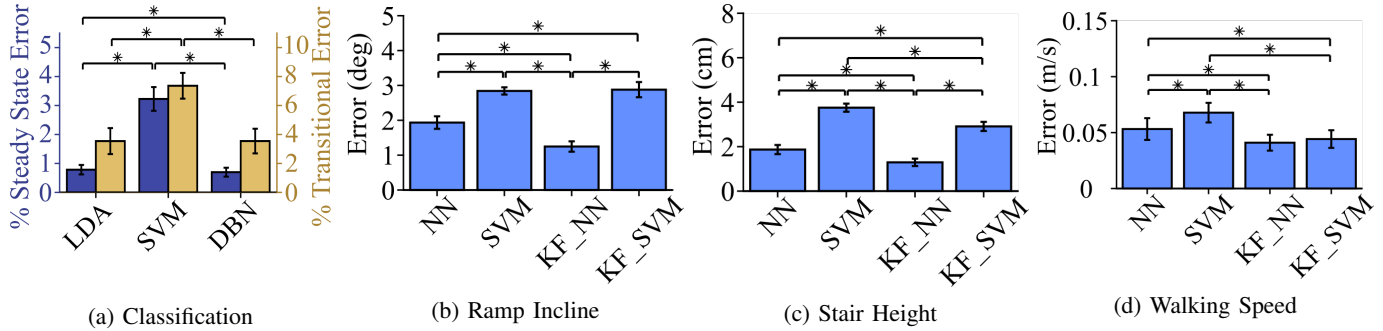


Fig. 5: Model selection for (a) classification and (b-d) regression. Error bars represent a 99% confidence interval, and a * indicates a significant change in error within subjects ($p < 0.01$). The SVM model performed worse than either of the other models, and the addition of time-history to classification (DBN) along with the Kalman filter for regression significantly decreased error within subjects compared to the LDA or plain NN.

C. Model Selection and Filtering

The results of the model selection process were similar between classification and all regressions (Figure 5). SVM models were found to perform significantly worse ($p < 0.01$) than all other models. The addition of time-history through the DBN and Kalman filter significantly reduced error compared to the initial models. For classification, the DBN was chosen for further analysis ($0.70 \pm 0.15\%$ steady state error, $3.55 \pm 0.85\%$ transitional error). For regression, the Kalman filtered NN was chosen for all ambulation modes, with an error of $1.25 \pm 0.15\text{deg}$ for ramps, $1.29 \pm 0.17\text{cm}$ for stairs, and $0.04 \pm 0.01\text{m/s}$ for walking speeds.

D. Model Optimization

Using the chosen DBN classifier, changing the number of classification points per gait cycle did not provide any significant improvement over the initial 4-point scheme, so the final errors were $0.70 \pm 0.15\%$ for steady state and $3.55 \pm 0.85\%$ for transitional. Likewise for the Kalman filtered NN, no significant improvements over the initial 1 layer of 20 nodes were found for estimating any of the parameters. This resulted in final errors of $1.25 \pm 0.15\text{deg}$, $1.29 \pm 0.17\text{cm}$, and $0.04 \pm 0.01\text{m/s}$. The full model optimization results can be found in the supplemental materials (Figure S2), along with confusion matrices (Figure S3) and other performance metrics for classification (Table S1).

E. Time-Series Representation

Although reducing the results to overall error numbers provides a concise summary, studying the model outputs over time can provide a better understanding of where errors occurred (Figure 6). For classification, Figure 6a shows the classifier predictions during 6 steps of the walk-stair descent-walk sequence, and corresponding figures for all four non-walking modes are provided in the supplemental materials (Figure S4). During a transitional step, the mode label deemed correct changes halfway through that step. Combining the transitional data for all modes, it was found that 88% of the transitional error prior to the label change was caused by the next mode being classified (early misclassification), and 92% of the transitional error after the label change was caused by the previous mode being classified (late misclassification). For regression, each estimator tracked the parameter accurately, with increases in error observed during changes in condition.

F. Sensor Importance

The importance of each sensor type along with the specific sensor locations within a type are shown in Figure 7. For classification, excluding IMU sensors resulted in the largest increase in error (122%) at steady state, followed by GON sensors (48%) and lastly EMG (29%). In the case of transitional error, excluding IMUs caused the largest increase in error followed by GON. Looking into forward selection of sensor locations for each type, the ankle GON, shank IMU,

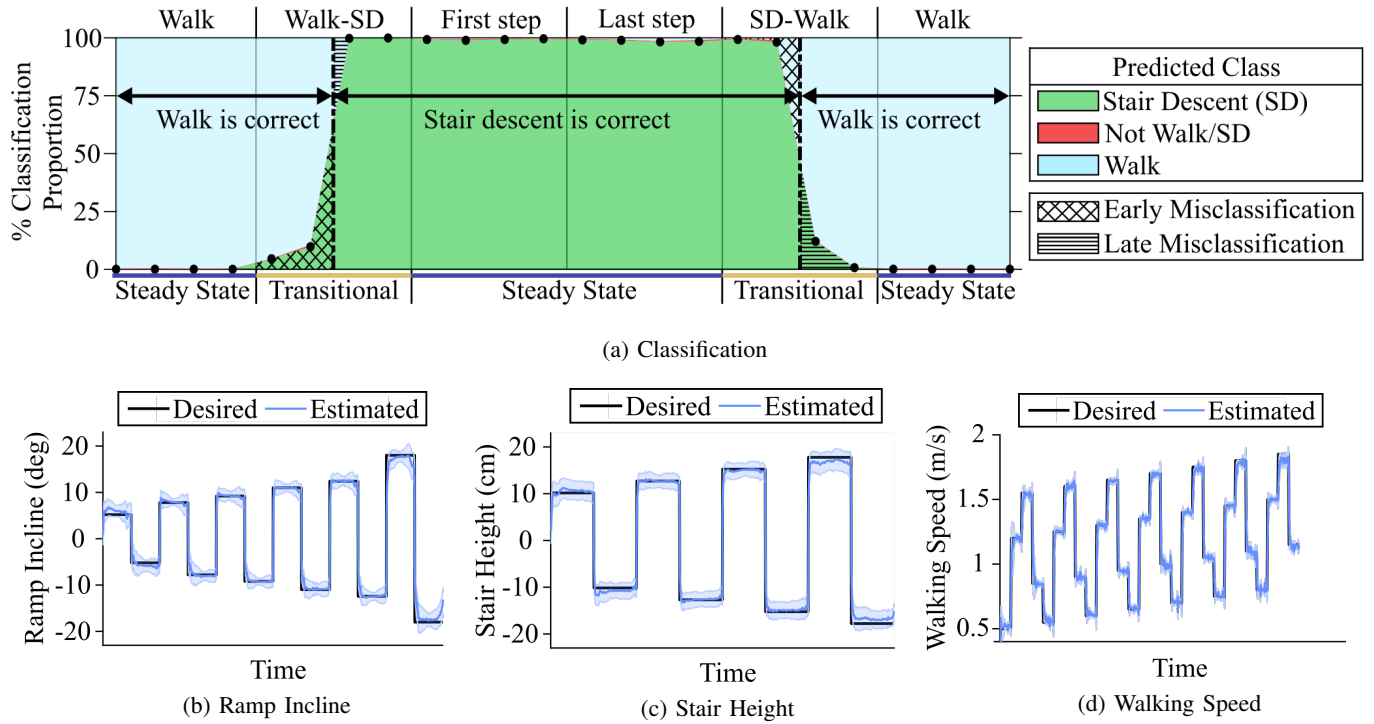


Fig. 6: Time series representation of (a) classification and (b-d) regression. For classification, the proportions of classifications during 6 steps of the walk-stair descent-walk sequence are shown. Steps were segmented toe-off to toe-off, and the step type (steady state or transitional) is listed below the x-axis. The mode label deemed correct during a transition changes halfway through the step, as indicated by dashed vertical lines. Corresponding figures for all four non-walking modes can be found in the supplemental materials. During transitional steps, most errors were caused by the next mode being classified before the label change (early misclassification) or the previous mode being classified after the label change (late misclassification). Combining transitional data from all modes, these early misclassifications accounted for 88% of errors before the label change, and late misclassifications accounted for 92% of errors after the label change. For regression, every estimator showed accurate tracking of the true condition, with larger error observed when the condition changes. All results are a subject average, and parameter estimation for the regression plots was normalized with respect to time. The shaded regions for regression represent ± 1 standard deviation.

and gastrocnemius EMG sensors were the most important of their respective sensor types. When estimating ramp incline, excluding GON sensors had a larger effect (183%) than excluding IMU (9%) and EMG (0%). Stair height estimation was also dependent on goniometers, as these sensors were much more important (280% increase in error) than IMU (14%) and EMG (0%). The only non-GON sensor selected was the shank IMU. Estimating walking speed was found to be heavily reliant on IMU sensors (475% increase) compared to excluding GON (1%), and EMG (0%). All IMUs were present in the selected set, however adding more than one IMU did not cause changes in error. The only non-IMU sensor chosen for speed estimation was the hip GON. EMG sensors were not present for any parameter estimation. Note that all percent changes presented in this section are relative to the “All” condition.

IV. DISCUSSION

This study illustrates the construction of a combined locomotion mode classifier and environmental parameter estimator to provide accurate information on a user’s current ambulation

state. Through the steps of feature selection, parameter sweeping, model selection, and model optimization, the proposed system achieved lower or similar error rates to the current state of the art for all four problems: mode classification, walking speed estimation, ground slope estimation, and stair height estimation. This is despite the evaluation of the proposed system over a much larger range of environmental parameters than previous studies, and thus being more robust to the full domain of possible locomotion. The feature dataset and Matlab code is provided to facilitate reproducibility and future work. Sensor importance analysis provides valuable information to designers as there were significant differences in the effects of each sensor type for these four problems.

For mode classification, the proposed system performed with greater than 98% accuracy on average (0.70% steady state error, 3.55% transitional error). These results are better than a recent able-bodied study by Hu et al. [19], which achieved 1.25% steady state error and 5.94% transitional error for an LDA utilizing unilateral EMG, GON, and IMU data. Aside from this error comparison, an important difference between the proposed system and previous literature is the

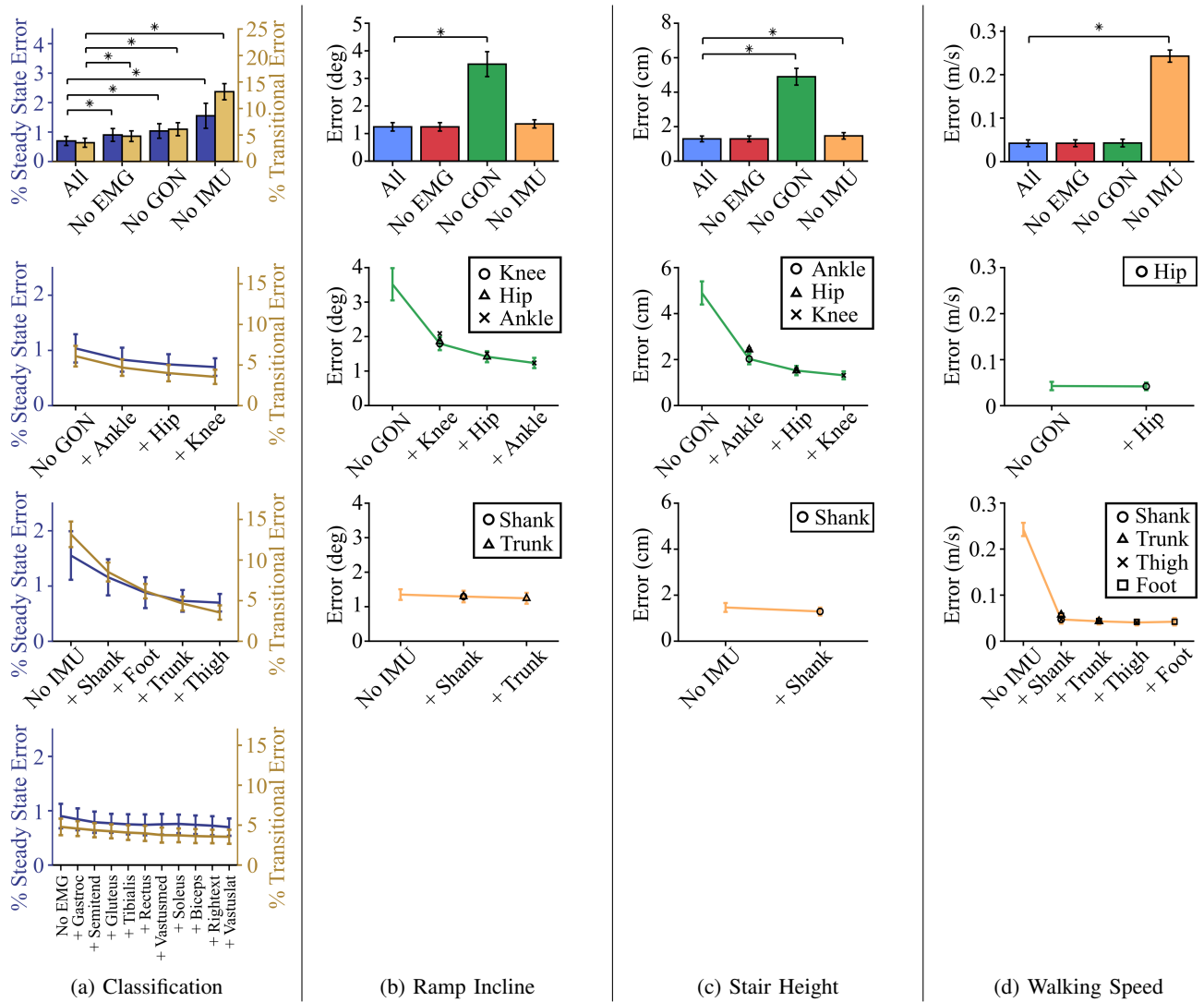


Fig. 7: Sensor importance for classification (a) and regression (b-d). Columns show the condition and rows show exclusion of sensor type on top followed by sensor location analysis for GON, IMU, and EMG sensors through forward sensor selection. Excluding any sensor type resulted in a significant increase in error for classification ($p < 0.01$; denoted by a *), while significant increases in error for regression only occurred when excluding GON for ramp incline, GON and IMU for stair height, and IMU for walking speed. It was also found that estimation of each parameter was dominated by a single sensor type. Markers in sensor selection plots for regression correspond to a sensor location. Error bars represent a 99% confidence interval, and no EMG sensors were present for any parameter estimation.

presence of multiple environmental parameter conditions for each mode. By leveraging the robust dataset, this classifier was able to identify a wide range of inclines and heights as the correct mode. Looking into the transitional behavior of our model (Figures 6a, S4), the increasing proportion of classifications of the next mode prior to the label change (early misclassifications) suggests that the model anticipated the mode change. Since steps were segmented by toe-offs, these early misclassifications occur during the swing phase going into the next mode. Counting these decisions as errors may therefore be inappropriate, as classifying the next mode during swing would allow for a device to achieve proper foot clearance and prepare for heel strike on the new mode. Excluding early misclassifications from transitional error resulted in an error of $1.64 \pm 0.51\%$.

Regarding parameter estimation, the root mean square error (RMSE) of estimation during steady state locomotion was computed. This allows for better comparison to previous studies which do not include transitions. For speed estimation, the proposed system's error of 0.07m/s RMSE was higher than both the 0.05m/s RMSE achieved through integration by Sabatini et al. [23] and the 0.03m/s RMSE resulting from a Gaussian Process Regression [25]. Both methods provided better performance but depend on algorithms that are more difficult to train and adapt due to the dependency on temporal parameters and increased computational complexity, respectively. Compared to the analysis from our research group of a neural network model with a similar sensor set however, the proposed system's performance was consistent with the published 0.08m/s RMSE [27]. The proposed system was also

able to estimate the larger range of speeds collected in this study, 0.5-1.85m/s compared to the 0.83-1.79m/s combined range of the other studies. These additional speeds at the low end are important as clinical populations tend to walk slower. This creates a harder regression problem since joint trajectories have been found to change for speeds slower than 0.69m/s [13]. For slope estimation, the steady state error of 1.30deg RMSE was the same as previous neural network models [27], however it was higher than the 0.87deg RMSE achieved through integration [23]. Similar to walking speed, the larger range of tested inclines, $\pm 18\text{deg}$ compared to a max of $\pm 10\text{deg}$ [27], is more challenging to estimate. This is especially important at the high end where subjects were observed to transition to toe striking. Since this is the first instance of estimating stair height, the steady state error of 1.5cm RMSE can serve as a benchmark for future studies.

Analysis of sensor importance showed that mechanical sensors (IMU, GON) were generally more important than EMG sensors. This was especially true for regression as no EMG features were selected for estimating any parameter. For classification, although the effect of removing EMG data was less pronounced than removing either mechanical sensor types, removing EMG still caused a significant increase in error. The hypothesis that regression would be dominated by mechanical sensors was strongly supported by sensor importance results, and this could be attributed to the nature of the parameter. Ramp incline and stair height define the geometry of the ambulation surface, and it was found that goniometers were crucial to their estimation. This suggests that the model was able to infer information about the surface geometry through the angular information of the lower limb joints. Speed estimation was performed almost entirely with IMUs, which aligns with previous methods for estimating speed. IMU sensor selection for speed also supported the common practice of using a single IMU, as most of the error was eliminated once the first sensor was selected. While the selection results showed the shank as the most important IMU location, this was a consequence of the forward selection algorithm. As shown by the markers at each selection step in Figure 7d, the differences in error between remaining IMU locations were typically smaller than the 99% confidence interval such that selecting any sensor location would greatly reduce the error. This was seen in all regression problems and indicates that the number of sensors matters more than their location for reducing error. As a whole, the sensor importance analysis showed that devices should emphasize the use of mechanical sensors over EMG; a favorable result for prostheses since EMG may be problematic depending on a patient's residual musculature. For devices aimed at parameter estimation, choosing the correct sensor type for the parameter is crucial, where goniometers were best for ramp incline and stair height estimation while IMUs should be used for speed estimation. The information found from this analysis can facilitate the implementation of real-time systems by informing the inclusion of only the most important sensors, thus reducing data bandwidth and computational load.

The optimization process produced interesting results as well. In the parameter sweeping step, it was found that longer windows were generally preferred for all conditions except

speed estimation. This trend agrees with previous literature [38], although local minima were only found for speed estimation and classification decisions at gait locations after 50% phase. This result for classification is notable as the preferred window size was split about the point where mode labels were changed. Although not shown in this study, this was consistent across DBN models with different numbers of classification points, suggesting that the point of label change influences the model characteristics. Another important result from the parameter sweep was the number of phase bins chosen for regression since this number determines the required resolution of the gait phase estimator. In the case of speed estimation, the choice of 2 phase bins meant that splitting the gait in two by detecting heel-strike and toe-off would suffice. Finally, the improvements in estimation due to Kalman-filtering the NN output strongly encourages the filtering of regression outputs.

This study was not without limitations and future areas of improvement. First, the analyses presented were performed offline. However, this machine learning strategy was developed with future online implementation in mind, such that similar models to those investigated have been previously implemented in real-time [22], [33]. Preliminary tests using an ODROID XU mini-PC have also been conducted to determine the latency of the system. These tests found that the strategy is feasible for real-time, with a total time for combined classification and regression under 20ms. Second, the strategy presented is based on able-bodied subjects only. This population was chosen so the results would generalize and not focus on a specific population. Many applications can benefit from this work, such as activity monitoring and human augmentation with exoskeletons. For use in clinical populations, implementing the proposed strategy would require further testing and optimization specific to the population.

Although the presented machine learning models performed well, there are other models and optimizations that may produce better results. We encourage others to use the published feature dataset for the analysis of other models, such as convolutional neural networks (CNNs) and Long Short Term Memory (LSTM) neural networks, along with thorough tuning of hyperparameters. The presented models were also only evaluated in a subject dependent manner, and the performance of subject independent models could be explored as well. One additional direction for future research is to apply the proposed method of parameter estimation to more environmental variables such as the user's heading or clinical measures like stride length, as these could allow for new device adaptation and improved rehabilitation.

V. CONCLUSION

Accurate knowledge of a user's current locomotion state is useful for many applications, such as allowing rehabilitative devices to adapt to changing environmental conditions or provide clinicians with more detailed information about a patient's activity. This study presents the development process of a combined classifier and parameter estimator that could be used for such purposes. The final models had high accuracy,

with the classifier achieving error rates of $0.70 \pm 0.15\%$ for steady-state and $3.55 \pm 0.85\%$ for transitions, while parameter estimation was performed with errors of $1.25 \pm 0.15\text{deg}$ for incline, $1.29 \pm 0.17\text{cm}$ for stair height, and $0.04 \pm 0.01\text{m/s}$ for speed. Analysis of which sensor types and sensor locations are important for each condition showed that mechanical sensors (IMU, GON) were most important for classification, while goniometers provided most of the accuracy for ramp/stair estimation and a single IMU was enough to reduce almost all the error in speed estimation. Future directions of the presented system are evaluating the performance for subjects with ambulatory disabilities and investigating the use of different models for classification and regression.

REFERENCES

- [1] E. Lauer and A. Houtenville, *Annual Disability Statistics Compendium: 2018*. Durham, NH: University of New Hampshire, Institute on Disability, 2018.
- [2] W. H. Organization, *World Report on Disability*, 2011, ch. 4.
- [3] F. Porciuncula, A. V. Roto, D. Kumar, I. Davis, S. Roy, C. J. Walsh, and L. N. Awad, "Wearable movement sensors for rehabilitation: A focused review of technological and clinical advances," *PM&R*, vol. 10, no. 9S2, pp. S220–S232, 2018.
- [4] H. Huang, F. Zhang, L. J. Hargrove, Z. Dou, D. R. Rogers, and K. B. Englehart, "Continuous locomotion-mode identification for prosthetic legs based on neuromuscular-mechanical fusion," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 10, pp. 2867–2875, Oct 2011.
- [5] P. Chinmilli, S. Redkar, W. Zhang, and T. Sugar, "A review on wearable inertial tracking based human gait analysis and control strategies of lower-limb exoskeletons," *Int Rob Auto J.*, vol. 3, no. 7, pp. 398–415, 2017.
- [6] F. Sup, A. Bohara, and M. Goldfarb, "Design and control of a powered knee and ankle prosthesis," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*, April 2007, pp. 4134–4139.
- [7] H. Varol, F. Sup, and M. Goldfarb, "Multiclass Real-Time Intent Recognition of a Powered Lower Limb Prosthesis," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 3, pp. 542–551, Mar. 2010.
- [8] K. Bhakta, J. Camargo, and A. J. Young, "Control and experimental validation of a powered knee and ankle prosthetic device," in *ASME 2018 Dynamic Systems and Control Conference*, vol. 1, 09 2018.
- [9] A. Simon, K. Ingraham, N. Fey, S. Finucane, R. Lipschutz, A. Young, and L. Hargrove, "Configuring a powered knee and ankle prosthesis for transfemoral amputees within five specific ambulation modes," *PloS one*, vol. 9, p. e99387, 06 2014.
- [10] W. Huo, S. Mohammed, J. C. Moreno, and Y. Amirat, "Lower Limb Wearable Robots for Assistance and Rehabilitation: A State of the Art," *IEEE Systems Journal*, vol. 10, no. 3, pp. 1068–1081, Sep. 2016.
- [11] W. Huo, S. Mohammed, Y. Amirat, and K. Kong, "Fast Gait Mode Detection and Assistive Torque Control of an Exoskeletal Robotic Orthosis for Walking Assistance," *IEEE Transactions on Robotics*, pp. 1–18, 2018.
- [12] Y. D. Li and E. T. Hsiao-Weckler, "Gait mode recognition and control for a portable-powered ankle-foot orthosis," in *2013 IEEE 13th International Conference on Rehabilitation Robotics (ICORR)*. Seattle, WA: IEEE, Jun. 2013, pp. 1–8.
- [13] H. van Hedel, L. Tomatis, and R. Müller, "Modulation of leg muscle activity and gait kinematics by walking speed and bodyweight unloading," *Gait & Posture*, vol. 24, no. 1, pp. 35 – 45, 2006.
- [14] J. T. Han, Y. H. Kwon, J. W. Park, H. M. Koo, and K. S. Nam, "Three-dimensional kinematic analysis during upslope walking with different inclinations by healthy adults," *Journal of Physical Therapy Science*, vol. 21, no. 4, pp. 385–391, 2009.
- [15] R. Riener, M. Rabuffetti, and C. Frigo, "Stair ascent and descent at different inclinations," *Gait & Posture*, vol. 15, no. 1, pp. 32 – 44, 2002.
- [16] S. Fritz and M. Lusardi, "White paper: 'walking speed: the sixth vital sign'," *J. Geriatr. Phys. Ther.*, vol. 32, pp. 2–5, 2009.
- [17] J. Lee and K. Seo, "The effects of stair walking training on the balance ability of chronic stroke patients," *J. Phys. Ther. Sci.*, vol. 26, pp. 517–520, 2014.
- [18] K. C. Seo and H. A. Kim, "The effects of ramp gait exercise with pnf on stroke patients' dynamic balance," *J. Phys. Ther. Sci.*, vol. 27, pp. 1747–1749, 2015.
- [19] B. Hu, E. Rouse, and L. Hargrove, "Fusion of bilateral lower-limb neuromechanical signals improves prediction of locomotor activities," *Frontiers in Robotics and AI*, vol. 5, p. 78, 2018. [Online]. Available: <https://www.frontiersin.org/article/10.3389/frobt.2018.00078>
- [20] H. Huang, T. A. Kuiken, and R. D. Lipschutz, "A strategy for identifying locomotion modes using surface electromyography," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 1, pp. 65–73, Jan 2009.
- [21] F. Attal, S. Mohammed, M. Dedabrishvili, F. Chamroukhi, L. Oukhellou, and Y. Amirat, "Physical human activity recognition using wearable sensors," Dec 2015.
- [22] A. J. Young, A. M. Simon, N. P. Fey, and L. J. Hargrove, "Intent Recognition in a Powered Lower Limb Prosthesis Using Time History Information," *Annals of Biomedical Engineering*, vol. 42, no. 3, pp. 631–641, Mar. 2014.
- [23] A. Sabatini, C. Martelloni, S. Scapellato, and F. Cavallo, "Assessment of Walking Features From Foot Inertial Sensing," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 3, pp. 486–494, Mar. 2005.
- [24] Jwu-Sheng Hu, Kuan-Chun Sun, and Chi-Yuan Cheng, "A Kinematic Human-Walking Model for the Normal-Gait-Speed Estimation Using Tri-Axial Acceleration Signals at Waist Location," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 8, pp. 2271–2279, Aug. 2013.
- [25] H. Vathsangam, A. Emken, D. Spruijt-Metz, and G. S. Sukhatme, "Toward free-living walking speed estimation using gaussian process-based regression with on-body accelerometers and gyroscopes," in *2010 4th International Conference on Pervasive Computing Technologies for Healthcare*, March 2010, pp. 1–8.
- [26] Z. He and W. Zhang, "Estimation of walking speed using accelerometer and artificial neural networks," in *Computer Science for Environmental Engineering and EcoInformatics*, Y. Yu, Z. Yu, and J. Zhao, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 42–47.
- [27] I. Kang, P. Kunapuli, H. Hsu, and A. J. Young, "Electromyography (emg) signal contributions in speed and slope estimation using robotic exoskeletons," in *2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR)*, June 2019, pp. 548–553.
- [28] G. Welch and G. Bishop, *An Introduction to the Kalman Filter*. University of North Carolina at Chapel Hill, 2006.
- [29] M. N. Alam, T. T. K. Munia, and R. Fazel-Rezai, "Gait speed estimation using kalman filtering on inertial measurement unit data," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, July 2017, pp. 2406–2409.
- [30] K. Deng, "Omega: On-line memory-based general purpose system classifier," Ph.D. dissertation, Carnegie Mellon University, Pittsburgh, PA 15213, 11 1998.
- [31] A. J. Young, T. A. Kuiken, and L. J. Hargrove, "Analysis of using EMG and mechanical sensors to enhance intent recognition in powered lower limb prostheses," *Journal of Neural Engineering*, vol. 11, no. 5, p. 056021, Oct. 2014.
- [32] "Differences in kinematics and electromyographic activity between men and women during the single-legged squat," *The American Journal of Sports Medicine*.
- [33] B. Su, J. Wang, S. Liu, M. Sheng, J. Jiang, and K. Xiang, "A cnn-based method for intent recognition using inertial measurement units and intelligent lower limb prosthesis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 5, pp. 1032–1042, May 2019.
- [34] J. Taborri, E. Palermo, S. Rossi, and P. Cappa, "Gait Partitioning Methods: A Systematic Review," *Sensors*, vol. 16, no. 1, p. 66, Jan. 2016.
- [35] F. Attal, Y. Amirat, A. Chibani, and S. Mohammed, "Automatic Recognition of Gait phases Using a Multiple Regression Hidden Markov Model," *IEEE/ASME Transactions on Mechatronics*, pp. 1–1, 2018.
- [36] P. Formento, R. Acevedo, S. Ghoussayni, and D. Ewins, "Gait Event Detection during Stair Walking Using a Rate Gyroscope," *Sensors*, vol. 14, no. 3, pp. 5470–5485, Mar. 2014.
- [37] J. Taborri, E. Scalona, E. Palermo, S. Rossi, and P. Cappa, "Validation of Inter-Subject Training for Hidden Markov Models Applied to Gait Phase Detection in Children with Cerebral Palsy," *Sensors*, vol. 15, no. 9, pp. 24 514–24 529, Sep. 2015.
- [38] A. J. Young, A. M. Simon, and L. J. Hargrove, "A training method for locomotion mode prediction using powered lower limb prostheses," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 3, pp. 671–677, May 2014.